

文章编号:1007-5321(2019)06-0170-07

DOI:10.13190/j.jbupt.2019-183

基于素描对视频的人脸识别研究

陈平¹, 皇甫大鹏¹, 王兴建^{1,2}, 党德鹏²

(1. 北京师范大学 信息网络中心, 北京 100875; 2. 北京师范大学 人工智能学院, 北京 100875)

摘要: 在公共安全领域查找关键人,需在视频中比对素描进行检索. 为此,改进了仅依靠对素描和视频数据中的人脸提取整体或者局部特征的识别算法,将人脸识别问题转为人脸检索问题,把这2种媒体数据表示为第3种数据,即图像列表. 通过比较图像列表完成最终的特征比对. 所设计的系统扩展了人脸识别的研究范围,且支持多种媒体数据的检索. 对比人脸视频分析融合系统,对素描-视频数据进行人脸识别,结果显示,所提出算法的正确识别率和曲线下面积都有相应提高,而等错误率降幅显著.

关键词: 素描; 视频; 人脸识别

中图分类号: TP391.4

文献标志码: A

Sketch Based Face Recognition in Video

CHEN Ping¹, HUANGFU Da-peng¹, WANG Xing-jian^{1,2}, DANG De-peng²

(1. Center of Information and Network Technology, Beijing Normal University, Beijing 100875, China;

2. School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China)

Abstract: For public securities, the sketch method is often need to be retrieved in video for seeking key people. An improved recognition algorithm that only extracts global or local features from sketch and video media data was presented. The problem of face recognition is turned into face retrieval. The above two medias data is represented as the so called third kind of data, i. e. image list and the final feature comparison is completed by comparing image list. The system designed here extends the research scope of face recognition and supports the retrieval of multi-media data. Compared with face recognition of face video analysis combined system for sketch-video data, the proposed design improves the correct recognition rate of precision and area under the curve, while the equal error rate of equal error rate decreases greatly.

Key words: sketch; video; face recognition

利用大量的监控视频数据和网络视频数据,可以通过检索图像视频,完成寻人、评估一个园区的人员流量、预测公众场合的突发事件、帮助警察抓捕嫌疑犯人等工作,应用十分广泛. 在抓捕犯人时应用的技术手段中,还包括对素描的检索. 在没有准确的身份信息,只有目击者供词的犯罪案件

时,就需要完成素描和图像的检索,以定位图像信息. 对于素描和视频这2种数据格式,其共同点是这二者都没有相对清晰、准确的图像特征,不同点在于素描描述的是静态图像信息,会因为画师的个人风格或者习惯对于同一个人有不同的素描呈现. 而视频描述的是动态图像数据,研究的数据大

收稿日期: 2019-07-24

基金项目: 国家自然科学基金项目(61370064, 61274033, 61271198); 赛尔网络下一代互联网技术创新项目(NGH20180603)

作者简介: 陈平(1974—),男,高级工程师,硕士生导师, E-mail: chenping@bnu.edu.cn.

多来自监控视频、互连网络、外部光线、个体运动,研究对象的大小、远近、清晰程度等因素都不可控制。正因为素描和视频数据有以上诸多限制因素,基于素描-视频的人脸识别研究显得尤为迫切,并具有很大的挑战性。

对于素描的人脸识别经历了从模拟合成到特征匹配的阶段,在模拟阶段,有2种方法识别素描和图像中的人脸:一种方法是将素描合成为图像,然后进行图像之间的比对;另一种方法是将图像合成素描,然后进行素描之间的比对。在此研究阶段,研究者无法直接利用素描或者图像本身的信息,也会因为在转化的过程丢失很多数据而影响匹配结果。随着在数字图像领域研究的深入,人们开始通过提取图像整体特征或者局部特征来比较图像和素描的特征,进行相似度之间的比较,一来充分利用原始信息,二来也提高了算法的可信度。Klare Brendan等^[1]运用局部特征判别分析方法(LFDA, local feature-based discriminant analysis),提取局部特征尺度不边特征变换(SIFT, scae-invariant feature transform)

算子特征和多尺度局部二进制模式(MLBP, multi-scale local binary patterns)特征;Galoogahi等^[2]使用局部拉东二分模式(LRBP, local Radon binary pattern)算法,将人脸形状转化到拉东(Radon)空间,在新空间中提取局部二分模式(LBP, local binary pattern)特征;Brendan Klared等^[3]提取SIFT算子,并结合素描-图像训练对的相似度;Liu J等^[4]运用联合字典的方法,解决异态(素描态和图像态)人脸识别问题;Yong Zhang等^[5]采用比较人眼识别人脸和主成分分析识别人脸。

1 素描-视频人脸识别算法

1.1 算法总体设计

在基于素描-视频的人脸识别过程中,对于人脸素描和人脸视频都要进行预处理、数据检索、特征比对这3个步骤。但视频由一系列帧图像构成,所有帧图像的特征信息构成了视频的特征信息,因此笔者设计了生成视频图像列表这一步骤。

整体算法流程如图1所示。

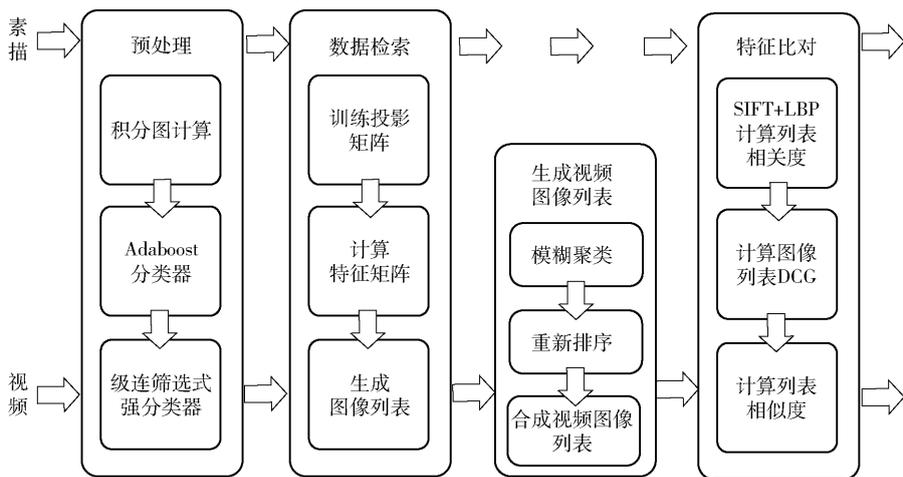


图1 素描-视频人脸识别整体算法

通过 Haar 特征分类器进行人脸检测,标准化人脸;其次通过 LDA 算法训练投影矩阵,计算图像第1级特征,并与数据库字典图像进行相似度比对,根据相似度对图像进行排序,生成图像列表;然后,通过第2级特征计算 SIFT + LBP 算子作为相关度,通过比较素描图像列表和视频图像列表计算相似度。

1.2 预处理

采用 Haar 特征分类器对图像进行人脸的检测^[6],检测的对象为素描图像和人脸帧图像(以下统称为图像)。根据 Haar 分类器检测出人脸

区域并对其进行标准化,以减少冗余信息,方便特征提取和分析^[7]。Haar 特征分类器的主要实现流程为^[8]:输入图像→对图像灰度进行积分图计算→AdaBoost 算法训练强分类器→级连筛选式强分类器→输出结果,其中 AdaBoost 算法是 Haar 分类器的核心算法。

1.2.1 AdaBoost

AdaBoost 是 Haar 分类器的训练系统,主要包括训练和选择 Haar 弱分类器生成强分类器,整体流程如图2所示。

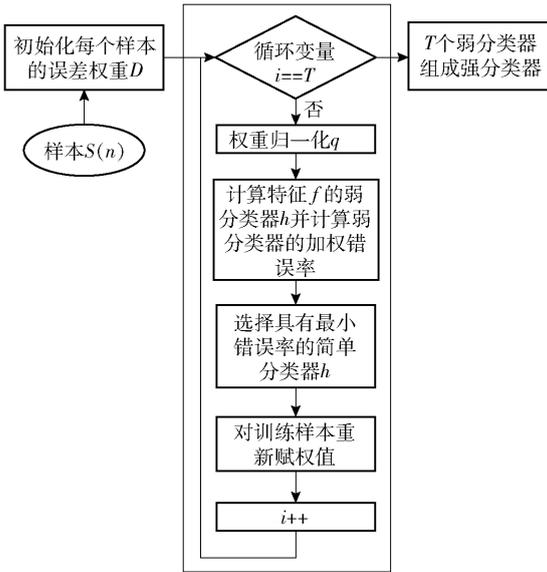


图 2 AdaBoost 算法流程

1.2.2 筛选式强分类器级连

筛选式强分类器级连是 Haar 的检测体系^[9],为了能检测出不同尺寸的人脸,提高算法适应范围,在 AdaBoost 算法的基础上做了改善. 强分类器的级连就是对现实世界包含多种尺寸人脸的图像进行多区域、多尺寸的检测即将图像划分多个区域,按照不同尺寸的窗口进行检测,最终输出包含人脸的子窗口图像.

如图 3 所示,筛选式强分类器的级连是通过调整每个强分类器中弱分类器的个数使得最终级连分类器的检测率 D 和误识率 F 达到系统值.

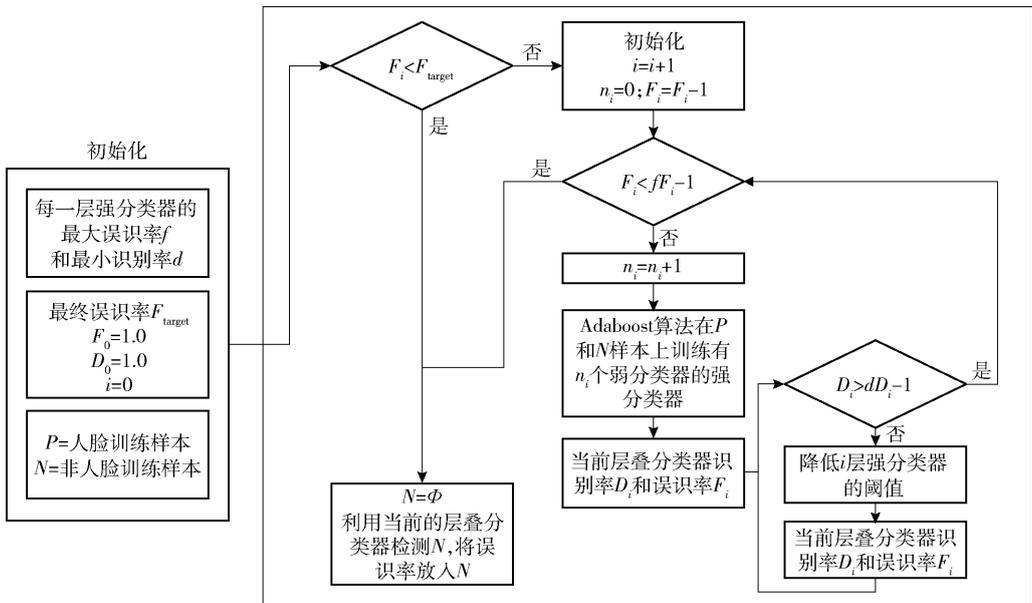


图 3 强分类器级连

1.3 数据检索

经过预处理过程获得标准人脸数据,然后进行图像数据检索. 判断两幅图像是否相似,比较它们的图像列表比直接比较它们自身的像素特征效果要好,进行数据检索最终是要生成人脸素描图像或人脸帧图像的图像列表;将人脸素描图像或人脸帧图像与数据库中的字典图像进行相似度比较,按照相似度排序生成的一系列图像作为人脸素描图像或人脸帧图像的图像列表. 数据检索的实现步骤如下:

步骤 1 选用局部尺度分析 (LDA, local de-mension analysis) 方法训练投影矩阵 W ,通过公示 $Y = W^T X$ 计算人脸素描图像或人脸帧图像的特征值 (其中 X 为输入图像, Y 为输出相应的特征矩阵);

步骤 2 通过欧式距离计算人脸素描图像或人脸帧图像与图像数据库中图像的相似度;

步骤 3 根据相似度由高到低进行排序,生成人脸素描图像或人脸帧图像的图像列表.

1.4 视频图像列表

对每一幅人脸帧图像都生成了图像列表,为了充分利用帧之间的信息,需整合每一幅帧图像的图像列表,生成视频的视频列表. 有如下假设.

1) 基于聚类(cluster)的假设. 对于与同一个查询,相关的文档倾向于聚在一起,因为它们与不相关的文档更相似. 文中对于这一假设的解释为:通过选择一个适当的聚类函数,可以把相关的文档从不相关的文档中分离出来,把相关的文档聚为一类(簇)或者几类,在类内包含的文档比其他文档更相

关,称这样的类为可信类。

2) 基于融合的假设. 基于相关文档检索到的不同排序列表之间重叠度很高,不相关文档的不同列表之间重叠度很低. 对于这一假设的解释为:基于融合的假设可以用来识别可信类,即一个类中包含的相关文档越多,那么这个类就越可信。

扩展到帧图像列表的处理,分3个小节描述:

1) 聚类. 对每一帧图像列表的字典图像特征矩阵进行聚类;2) 重新排序. 计算每一类类内相似度,并以此更新类内图像相似度,根据更新的相似度重新排序;3) 融合. 综合每一个字典图像在每个帧图像列表中的相似度,计算最终相似度,并依次排序生成视频的图像列表,如图4所示。

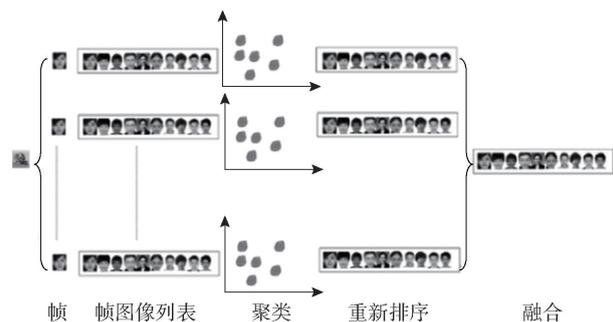


图4 帧图像列表生成过程

1.4.1 聚类

对每一帧图像列表中字典图像的特征矩阵值进行聚类分析。

$X = \{X_1, X_2, \dots, X_n\}$ 表示 n 个待聚类的目标,每个目标 X_i 包含 m 个属性,属性向量表示为 $[x_{i,1}, x_{i,2}, \dots, x_{i,m}]$,将 X 分成 k 个类,表示为 $C = [c_1, c_2, \dots, c_k]^T$, C 是 $k \times m$ 的矩阵,包含所有的聚类中心,所以模糊聚类算法(FCM, fuzzy c-means algorithm)的目标函数为

$$Ja(X, C, U) = \sum_{j=1}^k \sum_{i=1}^n (u_{i,j})^a D_{i,j}, \quad 1 \leq a < \infty \quad (1)$$

其中: a 为模糊加权指数,当 a 取值趋近 1 时,退化为 K -means 方法;取值趋向正无穷时, Ja 趋向于 0,根据文献[10], a 取 4。

U 是 $n \times k$ 的隶属度矩阵, $u_{i,j}$ 属于 U , 代表第 i 个元素属于类 c_j 的隶属度, $u_{i,j}$ 需要满足的限制条件为

$$\sum_{j=1}^k u_{i,j} = 1, u_{i,j} \in (0, 1], 1 \leq i \leq n \quad (2)$$

$D_{i,j}$ 代表计算欧式距离:

$$D_{i,j} = \sum_{l=1}^m (x_{i,l} - c_{j,l})^2, 1 \leq i \leq n, 1 \leq j \leq n \quad (3)$$

FCM 函数的目标是找到合适的分类 C 、隶属度矩阵 U ,使得 Ja 达到极小值,这样就可以保证经过 FCM 聚类算法的有效性. 为了计算 C 、 U ,在限制条件 $\sum_{j=1}^k u_{i,j} = 1, u_{i,j} \in (0, 1], 1 \leq i \leq n$ 的约束下,根据拉格朗日乘数法,计算

$$c_{j,l} = \frac{\sum_{i=1}^n (u_{i,j})^m x_{i,l}}{\sum_{i=1}^n (u_{i,j})^m}, 1 \leq j \leq k, 1 \leq l \leq m \quad (4)$$

$$u_{i,j} = \frac{1}{\sum_{l=1}^k \left(\frac{D_{i,j}}{D_{l,j}} \right)^{2/(m-1)}}, 1 \leq i \leq n, 1 \leq j \leq k \quad (5)$$

为了优化 FCM 的目标函数,只需要迭代计算上述 2 个公式,使得最终的隶属度矩阵相邻 2 次的变化收敛到设定的阈值 E ,此时 FCM 的目标函数达到极小. 实现流程如下:

步骤 1 初始化 C 、 U 、阈值 ε 、模糊加权指数 a 、聚类中心 V ;

步骤 2 通过式(4)和式(5)进行迭代:不断更新聚类中心 V 和隶属度矩阵 U ;

步骤 3 直到最终的聚类中心相邻两次的变化收敛到设定的阈值 ε 。

1.4.2 重排序

通过聚类算法之后,生成 C 类中的字典图像,重新计算其与帧图像相似度. 聚类是为了利用聚类中类内和类间的信息,对相似度进行重新计算. 因运用模糊聚类算法,图像是否属于某一类是由隶属度矩阵决定的. 为了公平地比较类之间的相似度,要保证每一类的图像个数是相等的,设为 count . 通过对类中元素隶属度的排序,选择该类中隶属度排名前 count 的字典图像组成类,其他依此类推. 每一类可靠性 R 的计算公式如下:

$$R(C_{l,j}) = \sum_{i=1, i \neq l}^n \sum_{p=1}^k \left[\frac{S_F C(F_i, C_{i,p})}{N_l} S(C_{l,j}, C_{i,p}) \right] \quad (6)$$

$$N_l = \sum_{i=1, i \neq l}^n \sum_{p=1}^k [S_F C(F_i, C_{i,p})] \quad (7)$$

$$S_F C(F_l, C_{l,j}) = \frac{\sum_{d \in C_{l,j}} \|F_l - d\|^2}{|C_{l,j}|} \quad (8)$$

$$S(C_{l,j}, C_{m,j}) = |C_{l,j} \cap C_{m,j}| \quad (9)$$

其中: d 为字典中的图像, norm_l 为图像列表中类的标准化因子, $|C_{l,j}|$ 为类 $C_{l,j}$ 中图像的数目, F_l 为视频的当前帧, $\|F_l - d\|^2$ 代表输入帧和字典图像的欧式距离; $S_F C(F_l, C_{l,j})$ 计算的是帧 F_l 和类 $C_{l,j}$ 的相似度; $S(C_{l,j}, C_{m,j})$ 计算的是 2 个类之间的相似度。

1.4.3 融合

对于每个字典图像, 经过聚类之后重新计算相似度, 这个相似度用于合成视频的全局图像列表。一段视频的每一帧图像之间相似度很大, 因而生成的图像列表中包含相同字典图像的比例很高, 为了

消除由此产生的冗余信息, 对分布在每一类中的同一个字典图像计算一个“终极”相似度, 计算公式为

$$S_d = \frac{\sum_{i=1}^n S_i^*(d)}{n} \quad (10)$$

根据相似度的值对字典图像进行排序, 融合为视频的图像列表。

1.5 特征比对

对每一幅人脸帧图像都生成了图像列表, 为了充分利用帧之间的信息, 需要整合每一幅帧图像的图像列表, 生成视频的图像列表, 具体流程如图 5 所示。

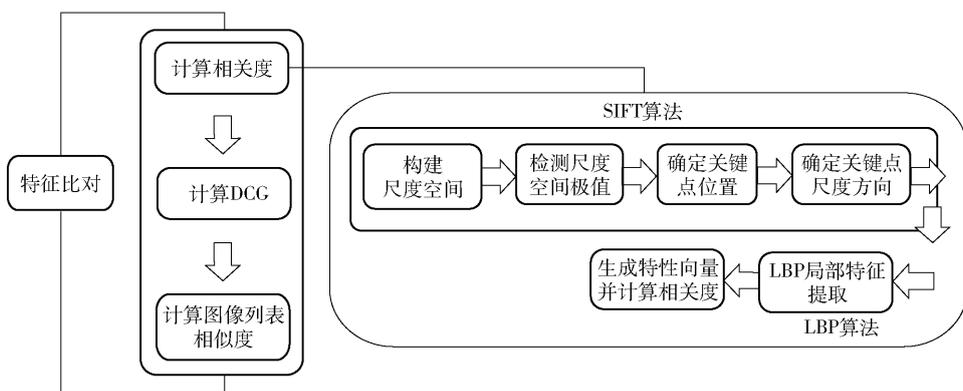


图 5 特征比对

1.5.1 相关度计算

先通过第 1 级特征计算人脸素描与字典图像、人脸视频与字典图像的相似度, 然后提取第 2 级局部特征, 结合 SIFT 算法和 LBP 算法, 计算特征直方图, 并使用 χ^2 距离来计算图像列表相关度 r 。

步骤 1 获取图像的关键点方向 θ 和尺度 σ 。根据 σ , 在相应尺度空间上, 以关键点为中心, 取 15×15 大小窗口; 按照 θ 将图像区域旋转到参考方向; 然后取 11×11 窗口大小为 LBP 遍历的窗口大小。

步骤 2 计算每一点的领域 LBP 特征值, 中心点的选择范围为 9×9 。对每一个中心点计算 LBP 特征值 (3×3), 距关键点越远的特征值对关键点的贡献越小, 对特征值进行加权, 加权系数由 LBP 中心点到关键点的距离确定。

步骤 3 根据 SIFT + LBP 生成图像的一维特征向量, 归一化之后计算相关度。

1.5.2 计算 DCG (discounted cumulative gain)

由图像列表中每一个字典图像的相关度, 计算整个图像列表的 DCG 值 D , 是衡量搜索引擎算法的

指标。在信息检索评价中, 相关度较高的文档排在检索列表较前位置被认为是好的检索结果。通过计算 D 来评价最终结果:

$$D[i] = \begin{cases} C[i], & \text{if } i = b \\ D[i-1] + G[i]/\log_b i, & \text{if } i \geq b \end{cases} \quad (11)$$

其中

$$C[i] = \begin{cases} G[1], & \text{if } i = 1 \\ C[i-1] + G[i], & \text{其他} \end{cases} \quad (12)$$

D 的计算公式如下 (q 为字典图像的排序):

$$D_q = \sum_{i=1}^{ \leq b} r_i + \sum_{i=b}^q \frac{r_i}{\log_b(i)} \quad (13)$$

1.5.3 相似度计算

消除样本数量对结果的干扰, 要对检索结果的 D 值进行标准化计算 (与理想状态下 IDCG 进行比较), 生成 nD 。应用在本文中, 如果 2 个图像列表 A 和 B 进行比对, A 把 B 作为 IDCG, 同时 B 把 A 作为 IDCG, 这样分别计算出 $nD(A, B)$ 和 $nD(B, A)$, 列表 A 和 B 最终的相似度由 $nD(A, B)$ 和 $nD(B, A)$ 的平均值确定。

以上的计算公式为

$$nD_q = \frac{D_q}{I_q} \quad (14)$$

$$S_i(S, V) = \frac{1}{2} \{ nD_q(S, V) + nD_q(V, S) \} \quad (15)$$

以上算法流程实现了素描—视频人脸识别的过程,先通过 Haar 特征分类器进行人脸检测,标准化人脸;紧接着通过 LDA 算法训练投影矩阵,计算图像第 1 级特征,并与数据库字典图像进行相似度比对,根据相似度对图像进行排序,生成图像列表;然后,通过第 2 级特征计算 SIFT + LBP 算子作为相关度,通过比较素描图像列表和视频图像列表 nD 计算相似度。

2 实验

对于人脸素描、人脸视频的特征提取、特征比对等算法均在 Matlab 环境下实现;对于图像的处理调用 Matlab 程序接口。

2.1 数据获取

实验需要用到的实验数据包括两类:相同的〈素描,视频〉数据和不同的〈素描,视频〉数据。相同的〈素描,视频〉数据指素描和视频代表的是同一个人;不同的〈素描,视频〉数据指素描和视频代表不同的人。本实验的数据来自互联网,数据举例如图 6 所示。

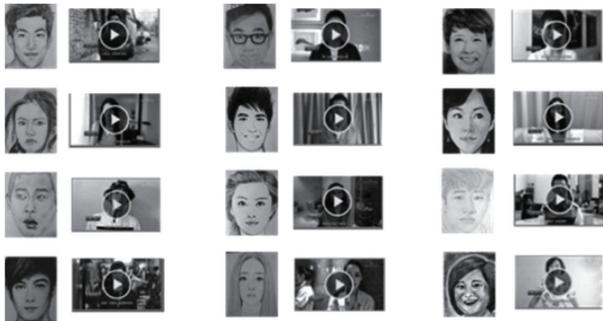


图 6 实验数据示例

2.2 素描—视频人脸识别实验

2.2.1 算法验证设计

实验中将不同类的〈素描,视频〉数据分为 10 组,每组含有 10 对〈素描,视频〉数据,即相同的〈素描,视频〉数据数量 = 不同的〈素描,视频〉数据数量 = $10 \times 10 = 100$ 个样本,样本涉及素描以及对应的一个视频,每个视频的场景各不相同,有街道监控类、课堂学习类、娱乐场所类、会场舞台类、在汽车交

通工具内等十余种场景。相同的〈素描,视频〉数据在采集数据时直接标注,不同的〈素描,视频〉数据随机配对。

验证算法有效性的过程就是计算人脸素描与人脸视频之间相似度的过程。应选择最相似的视频作为计算机算法获得的视频。对于相同的〈素描,视频〉数据,如果相似度最大视频刚好是实验数据中与人脸素描配对的视频,则算法正确识别;否则错误识别。对于不同的〈素描,视频〉数据,如果相似度最大视频刚好是实验数据中与人脸素描配对的视频,则算法错误接受;否则错误拒绝。

为了验证设计算法的有效性,设计对照实验:比较人脸视频分析融合系统(FaceVACS, face video analysis combined system)^[11]与本文算法在素描对于视频的人脸识别的效果。应用 FaceVACS 实现基于素描—视频人脸识别的方式:将素描与视频中的每一帧图像进行相似度比对,计算每一帧的平均相似度作为素描和视频之间的相似度。

2.2.2 对照实验结果分析

通过素描—视频人脸识别,评价指标选择正确识别率(P, precision)、等错误率(EER, equal error rate)和曲线下面积(AUC, area under the curve)、接受者操作特性曲线(ROC, receiver operating characteristic)。

从表 1 对比实验数据可以看出,对比 FaceVACS 对素描—视频数据进行人脸识别,所提出的设计在正确识别率和 AUC 2 个标准都有相应提高,而等错误率大大下降。

表 1 对照实验比较

算法	正确识别率	曲线下面积	等错误率
	P	AUC	ERR
FaceVACS	65.7	72.7	31.4
本文算法	80.3	89.6	18.7

图 7 为 ROC 曲线。从 ROC 曲线可以看出,随着伪阳性率逐渐增加,真阳性率呈现上升趋势,前面部分曲线上升比较迅速,后半部分上升比较缓慢,而在整体上升的趋势下,笔者提出的算法相比 FaceVACS 增长幅度更大。

3 结束语

实现了素描—视频人脸识别研究,通过素描图像检索出与之代表同一个人的视频信息,对于要检

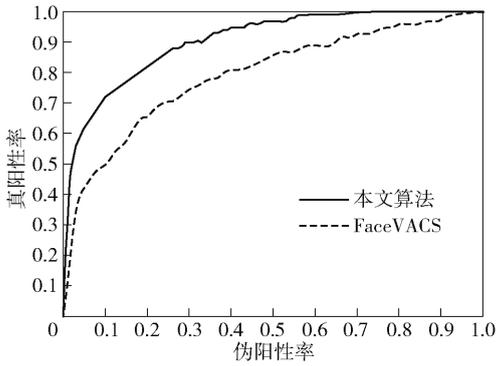


图7 ROC曲线

索的目标和被检索的对象,支持更多种数据格式,支持更多样化的检索任务,支持更复杂的检索需求,并从理论研究转为成果应用于实际. 后续工作主要包括算法优化、提高计算效率、与更多典型算法进行性能比较实验.

参考文献:

- [1] Klare Brendan, Zhifeng, Li A K, et al. Matching forensic sketches to mug shot photos[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(3): 639-646.
- [2] Kiani Galoogahi H, Sim T. Face sketch recognition by local radon binary pattern: LRPB[C]//2012 19th IEEE International Conference on Image Processing. New York: IEEE Press, 2012: 355-359.
- [3] Klare B, Jain A K. Sketch to photo matching: a feature-based approach[C]//Proceedings SPIE 7667, Biometric Technology for Human Identification VII. 2010(7667): 766702.
- [4] Liu J, Bae S, Park H, et al. Face photo-sketch recognition based on joint dictionary learning[C]//2015 14th IAPR International Conference on Machine Vision Applications (MVA). Tokyo: IEEE Press, 2015: 77-80.
- [5] Zhang Y, McCullough C, Sullins J R, et al. Hand-drawn face sketch recognition by humans and a PCA-based algorithm for forensic applications[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2010, 40(3): 475-485.
- [6] 龙伶俐. 基于 Adaboost 的人脸检测方法及眼睛定位算法研究[D]. 成都: 电子科技大学, 2008: 35.
- [7] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai: [s. n.], 2001: 311-318.
- [8] Viola Paul, Jones Michael. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [9] Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection[C]//Proceedings of International Conference on Image Processing. New York: IEEE Press, 2002: I900-I903.
- [10] Wu K L. Analysis of parameter selections for fuzzy cmeans[J]. Pattern Recognition, 2012, 45(1): 407-415.
- [11] Cognitec Systems GmbH. Face VACS-Video Scan[EB/OL]. 2019 [2019-04-16]. <https://www.cognitec.com/facevacs-videoscan.html>.